Greatly reduced amino acid alphabets in directed evolution: making the right choice for saturation mutagenesis at homologous enzyme positions†

Manfred T. Reetz* and Sheng Wu

Received (in Cambridge, UK) 1st August 2008, Accepted 12th September 2008
First published as an Advance Article on the web 2nd October 2008
DOI: 10.1039/b813388c

Enantioselective mutants of the thermally robust phenyl acetone monooxygenase (PAMO) as catalysts in Baeyer–Villiger reactions have been evolved by utilizing saturation mutagenesis in which drastically reduced amino acid alphabets are employed at homologous enzyme positions.

Directed evolution[1] of enantioselectivity and/or substrate acceptance of enzymes as catalysts in organic chemistry has emerged as a new approach to asymmetric catalysis.[2] Our contribution to making the method more efficient[3] is the combinatorial active-site saturation test (CAST),[4] in which the sites around the binding pocket of an enzyme are systematically considered for saturation mutagenesis with the formation of focused libraries[5] containing enantioselective variants, a given site being composed of one or more amino acid positions.[4] We have employed algorithms[6] to analyze oversampling[6,7] as a function of the % coverage of a library.[8] In a study involving an enzyme site comprised of three residues, it was shown that the quality of an NDT-library encoding a structurally balanced set of 12 amino acids is dramatically higher than that of a conventional NNK-library encoding the usual 20 amino acids (codon usage according to N: adenine/cytosine/guanine/thymine; K: guanine/thymine; D: adenine/guanine/thymine; T: thymine).[8b] Thus, considerably less screening for enantioselectivity is required.[8] However, in the case of a four residue site, the use of NNK or NDT codon degeneracy requires the screening of about 3.1 million or 162 000 transformants, respectively, for 95% coverage,[8] and when randomizing even larger regions, comprehensive screening turns into a hopeless task. In the present study we demonstrate how sequence alignment of homologous enzymes can help in making an appropriate choice of reduced amino acid alphabets,[9] thereby minimizing the screening effort.

We chose phenyl acetone monooxygenase (PAMO) from *Thermobifida fusca* as the enzyme to be engineered, which is the first and only thermostable Baeyer–Villiger monooxygenase (BVMO) identified thus far.[10,11] Unfortunately, PAMO has a narrow range of substrate acceptance, phenyl acetone and derivatives thereof being the only ketones that react with reasonable rates.[10] In a previous study based on rational design, two of the four residues in the loop spanning positions 441 to 444 next to the binding pocket were deleted.[12] They do not occur in the shorter loop of the thermally less stable CHMO which was known to accept a fairly wide variety of ketones,[13] and indeed a mutant with slightly broadened substrate scope was obtained.[12] In the present study we again considered the 441–444 loop, this time using a different strategy.

We first aligned the sequences[14] of WT PAMO and seven other Baeyer–Villigerases, namely STMO, CPMO, CDMO, CHMO, CHMO1, CHMO2 and CHMO3[13] (Fig. 1).

At positions 441, 442, 443 and 444 a limited number of amino acids appear, namely Ser/Ala, Ala/Val/Gly/Leu, Leu/Phe/Gly/Tyr and Ser/Ala/Cys/Thr, respectively. Rather than utilizing designed mixtures of oligonucleotides, we chose a more practical approach by applying appropriate codon degeneracies[6–8] which match as well as possible the amino acids occurring at the four positions while simultaneously introducing a limited number of additional amino acids for slightly higher structural diversity (Table 1).

As the model reaction we chose the oxidative kinetic resolution of *rac*-**1a**, which is hardly accepted by the WT PAMO. In a very slow reaction, only small amounts of product are formed after extended reaction times, the selectivity factor *E* amounting to 1.2 in slight favor of (*S*)-**2a**. Increasing the activity of the robust PAMO was thus the initial goal. CHMO from *Acinetobacter* TD3 is known to catalyze the reaction of *rac*-**1a** with an *E*-value of >100 in favor of (*R*)-**2a**,[15] but this BVMO is far less thermostable than PAMO. Other BVMOs have not been tested in these reactions thus far.

Using the codon degeneracies listed in Table 1, simultaneous randomization at all four positions in the 441–444 loop

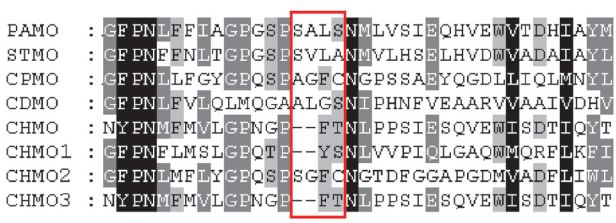

Fig. 1 Sequence alignment of BVMOs (441–444 loop in red box).

Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1, 45470 Mülheim/Ruhr, Germany.
E-mail: reetz@mpi-muelheim.mpg.de; Fax: +49 208 306 2985; Tel: +49 208 306 2000
† Electronic supplementary information (ESI) available: Experimental details. See DOI: 10.1039/b813388c

Table 1 Choice of codon degeneracies at each position in the 441–444 loop of PAMO. Degenerate codons: A (adenine); B (cytosine/guanine/thymine); C (cytosine); G (guanine); S (cytosine/guanine); for the definition of K and N, see text

Amino acid positions	Codon degeneracy	Encoded amino acids	Codons	Oversampling for 95% coverage
441	KCA	A, S		
442	KBG	S, A, L, V, W, G	864	2587
443	BGC	F, H, L, V, Y, G, D, R, C		
444	NSC	S, A, P, T, R, G, C		

of PAMO by saturation mutagenesis according to the QuikChange protocol[16] provided a library of PAMO variants which were subsequently evaluated as whole cell catalysts in the oxidative kinetic resolution of *rac*-**1**. For 95% coverage, only 2587 transformants would have to be screened.[8] We reduced the screening effort even more by arbitrarily restricting the search to 1700 clones, which were evaluated in the oxidative kinetic resolution of *rac*-**1a**. About 145 transformants were found to be more active than the WT PAMO in this transformation, which is a reflection of the high quality of the library.[8b] Had we striven for 95% coverage, even more hits would have been discovered. Seven of the best (*R*)-selective mutants and six of the best (*S*)-selective variants were sequenced (Table 2).

The results are striking in several respects. Whereas both (*S*)- and (*R*)-selective variants were identified, the latter showed considerably higher *E*-values, variant 254-21 displaying the highest enantioselectivity ($E = 70$ in favor of (*R*)-**2**). It is also noteworthy that such greatly reduced sets of amino acids, used as ''building blocks'' in the saturation mutagenesis experiment, lead to such different enantioselectivities. For example, the (*S*)-selective variant 121-10 is characterized by Ala/Leu/Val/Ser at positions 441/442/443/444, whereas the (*R*)-selective variant 254-67 with Ala/Ala/Asp/Ser differs at only two positions. Importantly, some of the few ''additional'' amino acids resulting from the choice of the codon degeneracies, beyond those occurring in the eight BVMOs as shown by the sequence alignment, appear in all of the best (*R*)-selective variants. Finally, some of the best mutants were isolated, purified and checked for thermostability relative to that of the robust WT PAMO. All of the variants tested (254-21, 254-60 and 254-67) showed half-lives at 50 °C comparable to that of the WT PAMO.[10,12] Thermostability is very high, *e.g.*, there is no loss of activity after a heat treatment at 50 °C for 40 h.

The gross mechanistic features of BVMOs are well known.[13] They are flavine-dependent enzymes in which molecular oxygen reacts with the reduced form of the flavine to generate an intermediate alkyl-hydroperoxide anion, which in turn adds nucleophilically to the carbonyl O-atom with formation of the short-lived Criegee-intermediate.[13] As evident from the X-ray structure of WT PAMO,[11] the 441–444 loop is near Arg337, the residue which has been postulated to stabilize the Criegee-intermediate.[11,12] The present results show that amino acid substitutions in the loop, and not just ''space-providing'' deletions as demonstrated previously,[12] influence substrate acceptance and enantioselectivity dramatically. Loops are generally considered to be flexible, making the interpretation of point mutations problematic,[1] but even in cases of reduced flexibility sound interpretations require detailed QM/MM studies, which we plan to perform for further insight.

We also studied the oxidative kinetic resolution of substrate *rac*-**1b** using some of the mutants identified earlier as catalysts for the conversion of *rac*-**1a**. Table 3 shows that enantioselectivity is even better in this case. The (*S*)-selective mutants in Table 2 were not tested because of their low enantioselectivity. Other ketones such as 2-[4-methylphenyl]cyclohexanone did not react, possibly indicating an electronic effect.

In conclusion, the primary message of this model study is the finding that enormously reduced amino acid alphabets can be used successfully in saturation mutagenesis at multi-residue sites, cases in which the conventional use of NNK codon degeneracy encoding all 20 amino acids would lead to a screening problem.[8b] This was illustrated by developing a simple strategy for the rapid directed evolution of reaction rate (substrate acceptance) and enantioselectivity of the thermally stable Baeyer–Villiger monooxygenase PAMO based on the use of reduced amino acid alphabets in saturation mutagenesis at all four positions of a loop. These alphabets were

Table 2 Oxidative kinetic resolution of *rac*-**1a** in whole-cell catalysis (WT sequence: Ser/Ala/Leu/Ser; mutations are marked boldface)

Mutant	Conversion (%)[a]	Enantioselectivity	*E*-Value	Sequence at positions 441–444
132-10	38	(*S*)	2.8	**Ala/Leu/Phe/Ala**
122-10	36	(*S*)	2.9	**Ala/Leu/Val/Ala**
132-18	45	(*S*)	1.8	**Ala/Ser/Phe/Ala**
121-02	41	(*S*)	1.4	Ser/**Ser/Phe/Gly**
121-10	32	(*S*)	1.5	**Ala/Leu/Val**/Ser
120-20	50	(*S*)	1.9	Ser/Ala/**Phe**/Ser
254-21	46	(*R*)	70	**Ala/Trp/Tyr/Thr**
122-02	48	(*R*)	31	Ser/**Trp/Arg**/Ser
254-58	15	(*R*)	48	Ser/**Trp/Tyr/Ala**
254-60	34	(*R*)	28	**Ala**/Ala/**Asp/Gly**
254-67	53	(*R*)	36	**Ala**/Ala/**Asp**/Ser
254-94	45	(*R*)	31	Ser/**Ser/Asp**/Ser
254-95	44	(*R*)	36	**Ala/Ser/Asp**/Ser

[a] Reaction time 24 h. Conditions: ESI.†

Table 3 Oxidative kinetic resolution of *rac*-**1b** with formation of (*R*)-**2b**

Variant	Conversion (%)[a]	*E*-value	Sequence at positions 441–444
254-21	59	>200	**Ala**/**Trp**/**Tyr**/**Thr**
254-58	36	>200	Ser/**Trp**/**Tyr**/**Ala**
254-60	39	147	**Ala**/Ala/**Asp**/**Gly**
254-67	51	187	**Ala**/Ala/**Asp**/Ser
254-94	39	133	Ser/**Ser**/**Asp**/Ser
254-95	48	43	**Ala**/**Ser**/**Asp**/Ser

[a] Reaction time 16 h. Conditions: ESI.†

defined by only two, six, nine and seven amino acids, respectively. The choice of the amino acids as building blocks was guided by sequence alignment of eight BVMOs focusing on a loop region, and supplemented by a select few additional amino acids as specified by appropriate degenerate codons. This protocol involved the screening of only 1700 transformants, yet a significant improvement of the catalytic profile was achieved, even though many other hits potentially accessible by conventional NNK codon usage are excluded by such a procedure. The pronounced robustness of PAMO was not compromised in the mutants, which means that these Baeyer–Villigerases may gain practical importance. We expect that this approach will find general application, specifically when randomizing whole loops or other large segments of enzymes or antibodies, *i.e.*, when the numbers problem in directed evolution[8b] becomes exceedingly acute.

We thank J. Rosentreter for chromatographic analyses and J. Zhou (group of Prof. B. List) for providing *rac*-**1b**.

Notes and references

1 Recent reviews of directed evolution: (*a*) K. M. Arndt and K. M. Müller, *Protein Engineering Protocols (Methods in Molecular Biology)*, Humana Press, Totowa, New Jersey, 2007, vol. 352; (*b*) F. H. Arnold and G. Georgiou, *Directed Enzyme Evolution: Screening and Selection Methods*, Humana Press, Totowa, New Jersey, 2003, vol. 230; (*c*) N. J. Turner, *Trends Biotechnol*., 2003, **21**, 474–478; (*d*) S. V. Taylor, P. Kast and D. Hilvert, *Angew. Chem*., 2001, **113**, 3408–3436 (*Angew. Chem., Int. Ed*., 2001, **40**, 3310–3335); (*e*) S. Brakmann and A. Schwienhorst, *Evolutionary Methods in Biotechnology (Clever Tricks for Directed Evolution)*, Wiley-VCH, Weinheim, 2004; (*f*) E. G. Hibbert, F. Baganz, H. C. Hailes, J. M. Ward, G. J. Lye, J. M. Woodley and P. A. Dalby, *Biomol. Eng*., 2005, **22**, 11–19; (*g*) S. B. Rubin-Pitel and H. Zhao, *Comb. Chem. High Throughput Screening*, 2006, **9**, 247–257; (*h*) J. Kaur and R. Sharma, *Crit. Rev. Biotechnol*., 2006, **26**, 165–199; (*i*) S. Bershtein and D. S. Tawfik, *Curr. Opin. Chem. Biol*., 2008, **12**, 151–158.

2 Reviews of directed evolution of enantioselective enzymes: (*a*) M. T. Reetz, *Proc. Natl. Acad. Sci. U. S. A*., 2004, **101**, 5716–5722; (*b*) M. T. Reetz, in *Advances in Catalysis*, ed. B. C. Gates and H. Knözinger, Elsevier, San Diego, 2006, vol. 49, pp. 1–69; (*c*) M. T. Reetz, in *Asymmetric Organic Synthesis with Enzymes*, ed. V. Gotor, I. Alfonso and E. García-Urdiales, Wiley-VCH, Weinheim, 2008, pp. 21–63.

3 See for example, (*a*) S. Lutz and W. M. Patrick, *Curr. Opin. Biotechnol*., 2004, **15**, 291–297; (*b*) R. J. Fox and G. W. Huisman, *Trends Biotechnol*., 2008, **26**, 132–138; (*c*) A. Herman and D. S. Tawfik, *Protein Eng., Des. Sel*., 2007, **20**, 219–226; (*d*) J. D. Bloom, M. M. Meyer, P. Meinhold, C. R. Otey, D. MacMillan and F. H. Arnold, *Curr. Opin. Struct. Biol*., 2005, **15**, 447–452; (*e*) T. S. Wong, D. Roccatano, M. Zacharias and U. Schwaneberg, *J. Mol. Biol*., 2006, **355**, 858–871.

4 (*a*) M. T. Reetz, M. Bocola, J. D. Carballeira, D. Zha and A. Vogel, *Angew. Chem*., 2005, **117**, 4264–4268 (*Angew. Chem., Int. Ed*., 2005, **44**, 4192–4196); (*b*) M. T. Reetz, L.-W. Wang and M. Bocola, *Angew. Chem*., 2006, **118**, 1258–1263 (*Angew. Chem., Int. Ed*., 2006, **45**, 1236–1241).

5 For examples of focused libraries,[1] see ESI†.

6 (*a*) W. M. Patrick and A. E. Firth, *Biomol. Eng*., 2005, **22**, 105–112; (*b*) A. E. Firth and W. M. Patrick, *Nucleic Acids Res*., 2008, **36**, W281–W285.

7 (*a*) L. Rui, Y. M. Kwon, A. Fishman, K. F. Reardon and T. W. Wood, *Appl. Environ. Microbiol*., 2004, **70**, 3246–3252; (*b*) A. D. Bosley and M. Ostermeier, *Biomol. Eng*., 2005, **22**, 57–61; (*c*) M. A. Mena and P. S. Daugherty, *Protein Eng., Des. Sel*., 2005, **18**, 559–561; (*d*) M. Denault and J. N. Pelletier, in *Protein Engineering Protocols*, ed. K. M. Arndt and K. M. Müller, Humana Press, Totowa, New Jersey, 2007, vol. 352, pp. 127–154.

8 (*a*) M. T. Reetz and J. D. Carballeira, *Nat. Protoc*., 2007, **2**, 891–903; (*b*) M. T. Reetz, D. Kahakeaw and R. Lohmer, *ChemBioChem*, 2008, **9**, 1797–1804.

9 Previous uses of reduced amino acid alphabets:[8b] (*a*) A. R. Davidson, K. J. Lumb and R. T. Sauer, *Nat. Struct. Biol*., 1995, **2**, 856–864; (*b*) M. H. Hecht, A. Das, A. Go, L. H. Bradley and Y. Wei, *Protein Sci*., 2004, **13**, 1711–1723; (*c*) K. U. Walter, K. Vamvaca and D. Hilvert, *J. Biol. Chem*., 2005, **280**, 37742–37746; (*d*) S. Akanuma, T. Kigawa and S. Yokoyama, *Proc. Natl. Acad. Sci. U. S. A*., 2002, **99**, 13549–13553; (*e*) A. D. Solis and S. Rackovsky, *Proteins: Struct., Funct., Bioinf*., 2007, **67**, 785–788; (*f*) N. Doi, K. Kakukawa, Y. Oishi and H. Yanagawa, *Protein Eng., Des. Sel*., 2005, **18**, 279–284; (*g*) J. Yang, Y. Koga, H. Nakano and T. Yamane, *Protein Eng*., 2002, **15**, 147–152; (*h*) C. M. Clouthier, M. M. Kayser and M. T. Reetz, *J. Org. Chem*., 2006, **71**, 8431–8437.

10 (*a*) M. W. Fraaije, J. Wu, D. P. H. M. Heuts, E. W. van Hellemond, J. H. L. Spelberg and D. B. Janssen, *Appl. Microbiol. Biotechnol*., 2005, **66**, 393–400; (*b*) G. de Gonzalo, D. E. Torres Pazmiño, G. Ottolina, M. W. Fraaije and G. Carrea, *Tetrahedron: Asymmetry*, 2005, **16**, 3077–3083; (*c*) D. E. Torres Pazmiño, B.-J. Baas, D. B. Janssen and M. W. Fraaije, *Biochemistry*, 2008, **47**, 4082–4093; (*d*) D. E. Torres Pazmiño, R. Snajdrova, B.-J. Baas, M. Ghobrial, M. D. Mihovilovic and M. W. Fraaije, *Angew. Chem*., 2008, **120**, 2307–2310 (*Angew. Chem., Int. Ed*., 2008, **47**, 2275–2278); (*e*) C. Rodríguez, G. de Gonzalo, M. W. Fraaije and V. Gotor, *Tetrahedron: Asymmetry*, 2007, **18**, 1338–1344.

11 E. Malito, A. Alfieri, M. W. Fraaije and A. Mattevi, *Proc. Natl. Acad. Sci. U. S. A*., 2004, **101**, 13157–13162.

12 M. Bocola, F. Schulz, F. Leca, A. Vogel, M. W. Fraaije and M. T. Reetz, *Adv. Synth. Catal*., 2005, **347**, 979–986.

13 Reviews of BVMOs: (*a*) C. T. Walsh and Y.-C. J. Chen, *Angew. Chem*., 1988, **100**, 342–352 (*Angew. Chem., Int. Ed. Engl*., 1988, **27**, 333–343); (*b*) J. D. Stewart, *Curr. Org. Chem*., 1998, **2**, 195–216; (*c*) S. Flitsch and G. Grogan, in *Enzyme Catalysis in Organic Synthesis*, ed. K. Drauz and H. Waldmann, Wiley-VCH, Weinheim, 2002, vol. 2, pp. 1202–1245; (*d*) V. Alphand, G. Carrea, R. Wohlgemuth, R. Furstoss and J. M. Woodley, *Trends Biotechnol*., 2003, **21**, 318–323; (*e*) P. C. Brzostowicz, D. M. Walters, S. M. Thomas, V. Nagarajan and P. E. Rouviere, *Appl. Environ. Microbiol*., 2003, **69**, 334–342; (*f*) M. D. Mihovilovic, F. Rudroff and B. Grotzl, *Curr. Org. Chem*., 2004, **8**, 1057–1069; (*g*) G.-J. ten Brink, I. W. C. E. Arends and R. A. Sheldon, *Chem. Rev*., 2004, **104**, 4105–4123; (*h*) R. Wohlgemuth, *Eng. Life Sci*., 2006, **6**, 577–583; (*i*) M. D. Mihovilovic, *Curr. Org. Chem*., 2006, **10**, 1265–1287.

14 Recent examples of the use of sequence data in mutagenesis:[1] (*a*) B. Steipe, B. Schiller, A. Plückthun and S. Steinbacher, *J. Mol. Biol*., 1994, **240**, 188–192; (*b*) N. Amin, A. D. Liu, S. Ramer, W. Aehle, D. Meijer, M. Metin, S. Wong, P. Gualfetti and V. Schellenberger, *Protein Eng., Des. Sel*., 2004, **17**, 787–793; (*c*) W. Besenmatter, P. Kast and D. Hilvert, *Proteins*, 2007, **66**, 500–506; (*d*) E. Vázquez-Figueroa, J. Chaparro-Riggers and A. S. Bommarius, *ChemBioChem*, 2007, **8**, 2295–2301.

15 V. Alphand, R. Furstoss, S. Pedrago-Moreau, S. M. Roberts and A. J. Willetts, *J. Chem. Soc., Perkin Trans. 1*, 1996, 1867–1872.

16 H. H. Hogrefe, J. Cline, G. L. Youngblood and R. M. Allen, *BioTechniques*, 2002, **33**, 1158–1165.